

P01-03185-US

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日      2003年  4月25日  
Date of Application:

出願番号      特願2003-122337  
Application Number:

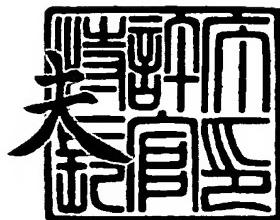
[ST. 10/C] :      [JP2003-122337]

出願人      パイオニア株式会社  
Applicant(s):

2003年12月26日

特許庁長官  
Commissioner,  
Japan Patent Office

今井康



【書類名】 特許願  
【整理番号】 58P0023  
【提出日】 平成15年 4月25日  
【あて先】 特許庁長官殿  
【国際特許分類】 G10L 3/00  
【発明者】  
【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社 総合研究所内  
【氏名】 小林 載  
【発明者】  
【住所又は居所】 埼玉県川越市山田字西町25番地1 パイオニア株式会社 川越工場内  
【氏名】 花井 建豪  
【特許出願人】  
【識別番号】 000005016  
【氏名又は名称】 パイオニア株式会社  
【代理人】  
【識別番号】 100083839  
【弁理士】  
【氏名又は名称】 石川 泰男  
【電話番号】 03-5443-8461  
【手数料の表示】  
【予納台帳番号】 007191  
【納付金額】 21,000円  
【提出物件の目録】  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【包括委任状番号】 9102133

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置、音声認識方法、並びに、音声認識プログラムおよびそれを記録した記録媒体

【特許請求の範囲】

【請求項 1】 発話された発話音声の音声成分を示す特徴量と予め格納されている発話が予想される音声の音声成分の特徴量を示す音声特徴量データとを照合させることによって当該発話音声を認識する音声認識装置であって、

複数の前記音声特徴量データが予め格納されている格納手段と、

前記格納されている音声特徴量データの種別毎に、予め設定された規則に基づいて、当該音声特徴量データを複数のデータ群に分類する分類手段と、

前記分類された各データ群毎に、前記データ群の特徴量を示すデータ群特徴量データを抽出する抽出手段と、

前記発話音声が発せられた環境の状況を示す環境データを取得する環境データ取得手段と、

前記格納されている各音声特徴量データ、前記分類された各データ群の属性を示す属性データ、前記取得された各データ群特徴量データ、および、前記環境データに基づいて、前記発話音声の特徴量に照合する際に用いられる前記音声特徴量データを生成する生成手段と、

前記生成された各音声特徴量データと前記発話音声の特徴量とを照合することによって前記発話音声を認識する認識手段と、

を備えたことを特徴とする音声認識装置。

【請求項 2】 請求項 1 に記載の音声認識装置において、  
前記抽出手段が、前記分類された各データ群毎に、前記データ群特徴量データとして前記各データ群における重心ベクトルのベクトルデータを抽出することを特徴とする音声認識装置。

【請求項 3】 請求項 1 または 2 に記載の音声認識装置において、  
前記生成手段が、  
前記各音声特徴量データと当該各音声特徴量データが属するデータ群における前記データ群特徴量データとの差異を示す差分特徴量を算出する第 1 算出手段と

前記取得された各データ群特微量データに前記環境データを、それぞれ、重畠して発話環境に適応したデータ群特微量データである適応データ群特微量データを算出する第2算出手段と、

前記算出された各音声特微量データにおける差分特微量および前記属性データと前記算出された各適応データ群特微量データに基づいて前記発話音声の特微量に照合する際に用いられる前記音声特微量データを生成する音声特微量データ生成手段と、

を有することを特徴とする音声認識装置。

**【請求項4】** 請求項3に記載の音声認識装置において、

前記第1算出手段および抽出手段の少なくとも何れか一方が、前記データ群特微量データまたは前記差分特微量を予め算出することを特徴とする音声認識装置。

**【請求項5】** 請求項3または4に記載の音声認識装置であって、

前記抽出手段が前記データ群特微量データとして前記各データ群における重心ベクトルのベクトルデータを抽出する場合に、

前記第1算出手段が、前記差分特微量として、前記各音声特微量データと当該各音声特微量データが属するデータ群における前記データ群特微量データとの差分ベクトルのベクトルデータを算出することを特徴とする音声認識装置。

**【請求項6】** 請求項1乃至5の何れか一項に記載の音声認識装置であって

前記発話された発話音声の特微量を認識すべきキーワードと当該キーワードを構成しない不要語とに識別することによって音声認識を行う場合に、

前記格納手段に前記キーワードの音声特微量データおよび前記不要語の音声特微量データが格納されているとともに、

前記分類手段が、前記キーワードおよび不要語の音声特微量データの種別毎に

予め設定された規則に基づいて、当該音声特微量データを複数のデータ群に分類することを特徴とする音声認識装置。

**【請求項 7】** 請求項 1 乃至 6 の何れか一項に記載の音声認識装置であって

前記発話音声を分析することによって当該発話音声の音声成分の特徴量である  
発話音声特徴量を取得する発話音声特徴量取得手段を備えるとともに、

前記認識手段が、

前記発話音声の少なくとも一部の音声区間の取得された前記発話音声特徴量と  
前記生成された音声特徴量データとを照合し、前記特徴量の特性が類似する割合  
を示す類似度を算出する類似度算出手段と、

前記算出した類似度に基づいて前記発話音声を認識する発話音声認識手段と、  
を有することを特徴とする音声認識装置。

**【請求項 8】** 発話された発話音声の音声成分を示す特徴量と予め格納され  
ている発話が予想される音声の音声成分の特徴量を示す音声特徴量データとを照  
合させることによって当該発話音声を認識する音声認識方法であって、

複数の前記音声特徴量データが予め格納されている格納手段から当該音声特徴  
量データを取得する音声特徴量データ取得行程と、

音声特徴量データの種別毎に、予め設定された規則に基づいて、当該音声特徴  
量データを複数のデータ群に分類する分類行程と、

前記分類された各データ群毎に、前記データ群の特徴量を示すデータ群特徴量  
データを抽出する抽出行程と、

前記発話音声が発せられた環境の状況を示す環境データを取得する環境データ  
取得行程と、

前記取得された各音声特徴量データ、前記分類された各データ群の属性を示す  
属性データ、前記取得された各データ群特徴量データ、および、前記環境データ  
に基づいて、前記発話音声の特徴量に照合する際に用いられる前記音声特徴量デ  
ータを生成する生成行程と、

前記生成された各音声特徴量データと前記発話音声の特徴量とを照合すること  
によって前記発話音声を認識する認識行程と、

を含むことを特徴とする音声認識方法。

**【請求項 9】** コンピュータによって、発話された発話音声の音声成分を示

す特微量と予め格納されている発話が予想される音声の音声成分の特微量を示す音声特微量データとを照合させることによって当該発話音声を認識する音声認識プログラムであって、

前記コンピュータを、

複数の前記音声特微量データが予め格納されている格納手段から当該音声特微量データを取得する音声特微量データ取得手段、

音声特微量データの種別毎に、予め設定された規則に基づいて、当該音声特微量データを複数のデータ群に分類する分類手段、

前記分類された各データ群毎に、前記データ群の特微量を示すデータ群特微量データを抽出する抽出手段、

前記発話音声が発せられた環境の状況を示す環境データを取得する環境データ取得手段、

前記取得された各音声特微量データ、前記分類された各データ群の属性を示す属性データ、前記取得された各データ群特微量データ、および、前記環境データに基づいて、前記発話音声の特微量に照合する際に用いられる前記音声特微量データを生成する生成手段、

前記生成された各音声特微量データと前記発話音声の特微量とを照合することによって前記発話音声を認識する認識手段、

として機能させることを特徴とする音声認識プログラム。

**【請求項 10】** 請求項 9 に記載の音声認識プログラムをコンピュータに読み取り可能に記録したことを特徴とする記録媒体。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、発話された音声からキーワードを認識する技術分野に属する。

##### 【0002】

##### 【従来の技術】

現在、人間が発声した音声を認識する音声認識装置が開発されており、このような音声認識装置では、人間が所定の語句の音声を発声すると、その入力信号か

ら語句の音声を認識するようになっている。

#### 【0003】

また、このような音声認識装置を車載されたナビゲーション装置やパソコン用コンピュータなど各種装置に適応すれば、その装置はキーボードやスイッチ選択等の手動操作を要することなく、各種の情報を入力することができるようになる。

#### 【0004】

したがって、自動車の運転中にナビゲーション装置を利用するなどの人間が両手を使用する作業環境であっても、操作者は、所望の情報を当該装置に入力することができるようになっている。

#### 【0005】

このような音声認識装置に用いられる代表的な音声認識方法にHMM（隠れマルコフモデル）と呼ばれる確率モデルを利用して音声認識を行う方法（以下、単に「音声認識」という）がある。

#### 【0006】

この音声認識は、発話音声の特徴量のパターンを、予め用意されたキーワードとなる認識候補の語句（以下、単に「キーワード」という）を示す音声の特徴量のパターンとマッチングさせることにより音声認識を行うようになっている。

#### 【0007】

具体的には、この音声認識は、入力された発話音声（入力信号）を予め定められた時間間隔毎に分析して特徴量を抽出し、この入力信号の特徴量に予めデータベースに格納されたHMMによって示されるキーワードとのマッチングする割合（以下、類似度という）を算出するとともに、発話音声の全てにおけるこの類似度を積算し、この積算された類似度が最も高いキーワードを認識結果として確定するようになっており、発話音声である入力信号から所定の語句の音声認識を行うことができるようになっている。

#### 【0008】

なお、HMMは、遷移する状態の集まりとして表される統計的信号源モデルであり、予めキーワードなどの認識すべき音声の特徴量を示す。また、このHMM

は、予め複数の音声データを採取し、これらの音声データに基づいて生成されるようになっている。

#### 【0009】

従来、このような音声認識では、発話音声に含まれるキーワード部分を如何に抽出するかが重要になる。

#### 【0010】

通常、発話音声には、キーワードの他に、予め既知の認識する際に不要な語である不要語（キーワードの前後に付加される「えー」や「です」等の語）が含まれるようになっており、発話音声は、原則的には、不要語と当該不要語に挟まれたキーワードによって形成される。したがって、音声認識を行う場合には、音声認識の対象となるキーワードを認識することによって行うワードスポットティングという手法（以下、単にワードスポットティング音声認識という）がよく用いられている。

#### 【0011】

このワードスポットティング音声認識は、認識対象となる発話音声を、キーワードモデルを示すHMMの他に、不要語のHMM（例えば、ガーベージモデル）を示すHMMを用意し、最も特徴量の類似度が高いキーワードのHMM、不要語のHMMまたはそれらの組み合わせを認識することによって音声認識を行うようになっている。

#### 【0012】

また、このワードスポットティング音声認識を用いた音声認識装置としては、雑音条件を考慮して音声認識を行うものが知られている（例えば、特許文献1参照）。

#### 【0013】

【特許文献1】 特開平10-161692号公報

#### 【0014】

【発明が解決しようとする課題】

しかしながら、上述の音声認識を用いた音声認識装置であっては、各HMMの特徴量は、学習用音声データを収録した条件（背景雑音、回線歪み、話者、声道

長など)に大きく依存し、この音声収録条件と実際の認識時の条件とが異なる場合には、入力音声データの特徴量とHMMの特徴量との不整合が生じ、結果として認識率が低下してしまう。

#### 【0015】

すなわち、HMMを生成するために、様々な発話音声データを収録するようになっており、当該収録した発話音声データに基づいて、当該HMMを生成するようになっているが、この音声収録条件と実際の認識時の条件とが異なると、各発話音声の特徴量も変わるために、入力データの特徴量とHMMの特徴量との不整合が生じ、結果として認識率が低下してしまう。

#### 【0016】

一方、入力データとHMMとの不整合による認識率の低下を防ぐには、認識を実行する際の条件と同じ条件で収録した音声データを使って、HMMを作成し直せばよいが、HMMのような統計的手法に基づくモデルは、膨大な量の学習音声データが必要で、処理に数百時間の時間がかかる。

#### 【0017】

本発明は、上記の各問題点に鑑みて為されたもので、その課題の一例としては、キーワードモデルおよび不要語モデル毎に、認識を実行する条件に基づいてこれらのモデルに対して適応処理を行い、音声認識の性能が高く、データ処理を迅速に行うことができる音声認識装置を提供することにある。

#### 【0018】

##### 【課題を解決するための手段】

上記の課題を解決するために、請求項1に記載の発明は、発話された発話音声の音声成分を示す特徴量と予め格納されている発話が予想される音声の音声成分の特徴量を示す音声特徴量データとを照合させることによって当該発話音声を認識する音声認識装置であって、複数の前記音声特徴量データが予め格納されている格納手段と、前記格納されている音声特徴量データの種別毎に、予め設定された規則に基づいて、当該音声特徴量データを複数のデータ群に分類する分類手段と、前記分類された各データ群毎に、前記データ群の特徴量を示すデータ群特徴量データを抽出する抽出手段と、前記発話音声が発せられた環境の状況を示す環

境データを取得する環境データ取得手段と、前記格納されている各音声特微量データ、前記分類された各データ群の属性を示す属性データ、前記取得された各データ群特微量データ、および、前記環境データに基づいて、前記発話音声の特微量に照合する際に用いられる前記音声特微量データを生成する生成手段と、前記生成された各音声特微量データと前記発話音声の特微量とを照合することによって前記発話音声を認識する認識手段と、を備えた構成を有している。

#### 【0019】

また、請求項8に記載の発明は、発話された発話音声の音声成分を示す特微量と予め格納されている発話が予想される音声の音声成分の特微量を示す音声特微量データとを照合させることによって当該発話音声を認識する音声認識方法であって、複数の前記音声特微量データが予め格納されている格納手段から当該音声特微量データを取得する音声特微量データ取得行程と、音声特微量データの種別毎に、予め設定された規則に基づいて、当該音声特微量データを複数のデータ群に分類する分類行程と、前記分類された各データ群毎に、前記データ群の特微量を示すデータ群特微量データを抽出する抽出行程と、前記発話音声が発せられた環境の状況を示す環境データを取得する環境データ取得行程と、前記取得された各音声特微量データ、前記分類された各データ群の属性を示す属性データ、前記取得された各データ群特微量データ、および、前記環境データに基づいて、前記発話音声の特微量に照合する際に用いられる前記音声特微量データを生成する生成行程と、前記生成された各音声特微量データと前記発話音声の特微量とを照合することによって前記発話音声を認識する認識行程と、を含む構成を有している。

#### 【0020】

また、請求項9に記載の発明は、コンピュータによって、発話された発話音声の音声成分を示す特微量と予め格納されている発話が予想される音声の音声成分の特微量を示す音声特微量データとを照合させることによって当該発話音声を認識する音声認識プログラムであって、前記コンピュータを、複数の前記音声特微量データが予め格納されている格納手段から当該音声特微量データを取得する音声特微量データ取得手段、音声特微量データの種別毎に、予め設定された規則に

基づいて、当該音声特微量データを複数のデータ群に分類する分類手段、前記分類された各データ群毎に、前記データ群の特微量を示すデータ群特微量データを抽出する抽出手段、前記発話音声が発せられた環境の状況を示す環境データを取得する環境データ取得手段、前記取得された各音声特微量データ、前記分類された各データ群の属性を示す属性データ、前記取得された各データ群特微量データ、および、前記環境データに基づいて、前記発話音声の特微量に照合する際に用いられる前記音声特微量データを生成する生成手段、前記生成された各音声特微量データと前記発話音声の特微量とを照合することによって前記発話音声を認識する認識手段、として機能させる構成を有している。

#### 【0021】

##### 【発明の実施の形態】

次に、本発明に好適な実施の形態について、図面に基づいて説明する。

#### 【0022】

なお、以下に説明する実施の形態は、本願に係る音声認識装置を車載用のナビゲーション装置に適応した場合の実施形態である。

#### 【0023】

また、本実施形態において説明する不要語構成要素は、音素単位または音節単位などの音声を構成する音声学的な基本単位を示すが、本実施形態では説明の便宜上、音節単位を用いて以下の説明を行う。

#### 【0024】

まず、図1を用いて本実施形態におけるナビゲーション装置の全体の構成および概要動作について説明する。

#### 【0025】

なお、図1は本願に係るナビゲーション装置の概要構成を示すブロック図である。

#### 【0026】

本実施形態のナビゲーション装置100は、図1に示すように、図示しないアンテナATに接続され、GPS (Global Positioning System) データを受信するGPS受信部110と、車両の走行速度などの走行データを検出するセンサ

部120と、GPSデータおよび走行データに基づいて自車位置を算出するインターフェース130と、VICS(Vehicle Information Communication System)データを受信するVICSデータ受信部140と、地図データが予め格納されている地図データ格納部150と、ユーザが各設定を行うとともにシステムに命令を入力する際に用いられる操作部160と、ユーザの音声を集音するマイクロホン170と、マイクロホン170によって集音された音声からシステムに指示された命令を認識する音声認識部300と、地図データおよび車両の位置を表示する表示部180と、バッファメモリ190を用いて表示部180を制御する表示制御部200と、経路誘導などの音声を生成する音声処理回路210と、音声処理回路210から出力された音声信号を拡声するスピーカ220と、電話公衆網回線やインターネット回線とアンテナATを介して通信を行う通信部230と、経路探索などのナビゲーション処理を行うとともに、システム全体の制御を行うシステム制御部240と、RAM/ROM250と、により構成され、システム制御部240と各部は、バス260により接続されている。

#### 【0027】

GPS受信部110は、GPSに属する複数の人工衛星からの航法電波を、アンテナATを介して受信するとともに、この受信した電波に基づいて移動体の現在位置の擬似座標値を計算してGPSデータとしてインターフェース130に出力するようになっている。

#### 【0028】

センサ部120は、車両の走行速度、加速度および方位角の走行データを検出するようになっており、検出した走行データをインターフェース130に出力するようになっている。

#### 【0029】

具体的には、センサ部120は、車両の走行速度を検出し、その検出した速度をパルス又は電圧の形態を有する速度データに変換してインターフェース130に出力するようになっている。

#### 【0030】

また、センサ部120は、重力加速度と車両の移動により発生する加速度とを

比較することにより、上下方向の車両の移動状態を検出し、当該検出した移動状態を示す加速度データをパルス又は電圧の形態に変換してインターフェース130に出力するようになっている。

#### 【0031】

さらに、センサ部120は、いわゆるジャイロセンサにより構成され、車両の方位角、即ち車両が進行している進行方向を検出し、検出した方位角をパルス又は電圧の形態を有する方位角データに変換してインターフェース130に出力するようになっている。

#### 【0032】

インターフェース130は、センサ部120およびGPS受信部110とシステム制御部240との間のインターフェース処理を行うようになっており、入力されたGPSデータと走行データに基づいて自車位置を算出して当該自車位置を自車位置データとしてシステム制御部240に出力するようになっている。

#### 【0033】

なお、この自車位置データは、システム制御部240において地図データと照合されてマップマッチング処理等が実行されるようになっている。

#### 【0034】

VICSデータ受信部140は、FM多重放送などの電波を受信することによってVICSデータを取得するようになっており、取得したVICSデータをシステム制御部240に出力するようになっている。

#### 【0035】

なお、VICSとは、道路交通情報通信システムのことを示し、VICSデータとは、渋滞、事故、規制などの道路交通情報をいう。

#### 【0036】

地図データ格納部150は、例えば、ハードディスク(Hard Disc)によって構成されるようになっており、予め記録されている道路地図などの地図データその他の走行案内に必要な情報(以下、地図データ等という。)を読み出すとともに、読み出した地図データ等をシステム制御部240に出力するようになっている。

**【0037】**

この地図データ等とは、ナビゲーション動作に必要な道路形状データを含む地図データの他、公園や店舗などの目的地の名称、位置データなどの各種関連データが道路形状データに対応付けられて記憶されているとともに、地図全体がメッシュ状の複数のブロックに分割され、各ブロックに対応する地図データがブロック地図データとして管理されるようになっている。

**【0038】**

操作部160は、各種確認ボタン、選択ボタン及び数字キー等の多数のキーを含むリモートコントロール装置等により構成されており、特に、音声認識の処理開始命令などの運転者の命令を入力するために用いられるようになっている。

**【0039】**

音声認識部300には、マイクロホン170に入力されたユーザから発生された発話音声が入力されるようになっており、例えば、当該音声認識部300は、ナビゲーション装置100の操作コマンドとして入力された発話音声を解析し、解析結果をシステム制御部240に出力するようになっている。

**【0040】**

なお、本実施形態における音声認識部300の構成および動作の詳細については後述する。

**【0041】**

表示部180は、例えば、C R Tまたは液晶表示素子によって構成され、表示制御部200の制御にしたがって地図データ等を種々の態様で表示するとともに、これに重畠して自車位置などの経路案内に必要な各種状態を表示するようになっている。

**【0042】**

なお、この表示部180は、地図データ等以外のコンテンツ情報を表示するようになっており、表示制御を介してシステム制御部240の指示に基づいて各コンテンツ情報を表示するようになっている。

**【0043】**

表示制御部200には、システム制御部240を介して入力された地図データ

等が入力されるようになっており、表示制御部200は、このシステム制御部240の指示に基づいて上述のような表示部180に表示すべき表示データを生成し、バッファメモリ190に一時的に保存しつつ、所定のタイミングでバッファメモリ190から表示データを読み出して表示部180に表示出力するようになっている。

#### 【0044】

音声処理回路210は、システム制御部240の指示に基づいて音声信号を生成し、生成した音声信号を、スピーカ220を介して拡声するようになっており、例えば、次の交差点における車両の進行方向や走行案内上運転者に直接告知すべき渋滞情報又は通行止め情報等を含む経路誘導に関する情報、および、音声認識部300において認識された音声認識結果を音声信号としてスピーカ220に出力するようになっている。

#### 【0045】

システム制御部240は、GPS受信ポート、キー入力ポート、表示制御ポート等の各種入出力ポートを含み、ナビゲーション処理のための全般的な機能を総括的に制御するようになっている。

#### 【0046】

このシステム制御部240は、ナビゲーション装置100全体の動作を制御するようになっており、RAM/ROM250に格納される制御プログラムを読み出して各処理を実行するとともに、当該RAM/ROM250に処理中のデータを一時的に保持することによって経路誘導のための制御を行うようになっている。

#### 【0047】

例えば、このシステム制御部240は、ナビゲーションの処理を行う場合に、音声認識部300、特に後述する適応処理部330を制御して、マイクロホン170から入力されたユーザの発話音声を解析させるとともに当該発話音声を認識させるようになっており、この認識させた発話音声からナビゲーション処理の命令コマンドを取得し各部の制御を行うようになっている。

#### 【0048】

次に、図2を用いて本実施形態の音声認識部の構成および動作について説明する。

#### 【0049】

なお、図2は、本実施形態における音声認識部の構成を示すブロック図である。

#### 【0050】

本実施形態の音声認識部300は、図2に示すように、キーワードモデルおよび不要語モデルを予め格納するHMMデータベース310と、マイクロホン170を介して入力されたマイク入力信号に基づいて雑音モデルと雑音特徴量とを算出する雑音推定処理部320と、システム制御部240の指示により雑音モデルに基づいてデータベースに格納されている各HMMに対して適応処理を行う適応処理部330と、入力されたマイク入力信号全体の雑音低減処理を行う雑音低減処理部350と、マイク入力信号をフレーム毎に分析してマイク入力信号の特徴量を抽出する音声分析部360と、音声分析された各フレーム毎の特徴量とHMMデータベース310に格納されている各HMMの特徴量とを比較し、各フレーム毎の類似度を算出する類似度算出部370と、入力されたマイク入力信号により、音声区間の推定を行う音声区間推定部380と、音声区間の情報と各フレーム毎の出力確率および遷移出力確率とに基づいて各キーワードモデルと不要語モデルとの各組み合わせの累積類似度を各キーワード毎に算出するマッチング処理部390と、各キーワード毎に算出された累積類似度の中で最も累積類似度が高いキーワードをキーワードと判定する判定部400と、を備えている。

#### 【0051】

また、雑音推定処理部320、雑音低減処理部350および音声区間推定部380は、バス260を介して発話音声信号が入力されるとともに、判定部400は、バスを介して判定したキーワードをシステム制御部240に出力するようになっている。

#### 【0052】

なお、例えば、本実施形態のHMMデータベース310は、本発明の格納手段を構成し、雑音推定処理部320は、本発明の環境データ取得手段を構成する。

**【0053】**

また、例えば、本実施形態の適応処理部330は、本発明の分類手段、抽出手段、生成手段および音声特徴量データ取得手段を構成し、音声分析部360は、本発明の発話音声特徴量取得手段を構成する。

**【0054】**

さらに、例えば、本実施形態の類似度算出部370は、本発明の認識手段および類似度算出部370を構成し、マッチング処理部390および判定部400は、本発明の認識手段および発話音声認識手段を構成する。

**【0055】**

HMMデータベース310には、認識すべき複数のキーワードの特徴量のパターンデータを示すキーワードのHMM（以下、単に「キーワードモデル」という。）と、キーワード以外の言語である不要語における特徴量のパターンデータを示すHMM（以下、単に「不要語モデル」という。）が格納されており、このHMMデータベース310は、格納してある各HMMモデルを適応処理部330に出力するようになっている。

**【0056】**

具体的には、HMMデータベース310には、格納してある各HMMモデルに対して、ユーザ毎に、すなわち、話者毎に適応処理が為された各HMMが格納されており、この当該話者毎に設定された各HMMを適応処理部330に出力するようになっている。

**【0057】**

例えば、システム制御部240は、予め設定された文章をユーザに読ませ、マイクロホン170を介して入力されたその文章の発話音声特徴量に基づいて音声認識処理を行う前に、HMMデータベースに格納されているHMMの適応処理を行うようになっており、音声認識の処理を行う毎に当該話者適応された各HMMを、HMMデータベース310に格納しておくようになっている。

**【0058】**

また、本実施形態では、HMMデータベース310には、キーワードモデルとしては、自動車が向かう目的地名や現在位置名、レストランなどの施設名といっ

た音声信号の特徴量のパターンを示すHMM（以下、「キーワードモデル」という。）が格納されるようになっている。

#### 【0059】

なお、通常、キーワードは、「現在地」や「目的地」のように、複数の音節または音素から構成され、1つのキーワードモデルは複数のキーワードを構成するHMMによって構成されている。したがって、類似度算出部370では、各キーワードの構成要素となるHMM毎に1のフレーム毎の特徴量における類似度を算出するようになっているため、実際には、HMMデータベース310には、各キーワードモデル毎に、当該キーワードを構成するHMMが格納されている。

#### 【0060】

さらに、本実施形態では、HMMデータベース310には、複数の不要語の特徴量を示すHMMが（以下、「不要語モデル」という。）が格納されており、例えば、音素毎、または、音節毎に、纏められた複数の音声特徴量のパターンが不要語構成モデルとして格納されている。

#### 【0061】

なお、各音節毎にそれぞれの特徴量のHMMを生成する場合には、上述したように、予め複数の人間が発する各音節の音声データを取得するとともに、各音節毎に特徴量のパターンを抽出し、当該各音節毎の特徴量のパターンに基づいて各音節の特徴量のパターンデータを学習させて生成するようになっている。

#### 【0062】

また、本実施形態では、不要語モデルの音声データを生成する際に、複数の音節の音声データに基づいて1の特徴量パターンのHMMを生成し、1のHMMの言語モデルによって複数の音節の特徴量を有するHMMを音韻毎などの予め複数生成した不要語モデルをHMMデータベース310に格納するようになっている。

#### 【0063】

雑音推定処理部320には、マイクロホン170を介してユーザ発話を含む入力信号（以下、「マイク入力信号」という。）が入力されるようになっており、この雑音推定処理部320は、入力されたマイク入力信号から予め設定された初

期区間のマイク入力信号（以下、「初期入力信号」という。）に基づいて雑音モデルを算出するとともに、入力されたマイク入力信号に基づいて雑音低減処理のための雑音音声の特徴量（以下、「雑音特徴量」という。）を算出するようになっている。

#### 【0064】

具体的には、雑音推定処理部320は、システム制御部240によって検出されたユーザの発話開始に基づいて、当該ユーザの発話が開始された時刻から予め設定された区間のマイク入力信号を初期入力信号として取得するようになっており、この取得した初期入力信号に基づいて、後述する適応処理のための雑音モデルを算出し、当該算出した雑音モデルをデータとして適応処理部330に出力するようになっている。

#### 【0065】

なお、本実施形態では、例えば、この雑音推定処理部320は、マイク入力信号の開始時刻から数百msec程度の区間のマイク入力信号を取得するようになっている。また、この雑音モデルとは、後述する適応処理のための変数（パラメータ）の一つであり、マルケプストラムまたはLPC（Linear Predictive Coding）ケプストラムに代表される特徴パラメータとして表現される。

#### 【0066】

また、雑音推定処理部320は、マイク入力信号に基づいて、当該マイク入力信号の一定時間毎の各周波数毎におけるパワーを示すスペクトル包絡、または、このパワースペクトルの対数をとって逆フーリエ変換して算出されたケプストラムの特徴量（以下、「雑音特徴量」という）をデータとして雑音低減処理部350に出力するようになっている。

#### 【0067】

適応処理部330には、雑音推定処理部320によって算出された雑音モデルのデータ（以下、単に、「雑音モデル」という。）が入力されるようになっており、この適応処理部330は、入力された雑音モデルに基づいて、HMMデータベース310に格納されているキーワードのHMMのデータであるキーワードモデルおよび不要語のHMMのデータである不要語モデルにおける各HMMのデータ

タ適応処理を行うようになっている。

#### 【0068】

具体的には、適応処理部330は、各キーワードモデルおよび不要語モデル毎に予めクラスタリング処理およびその他の処理を行うとともに、クラスタリング処理等の結果に基づいて、各HMMのデータ適応の処理（以下、「データ適応処理」という。）を行うようになっており、データ適応処理をした各HMMを類似度算出部370に出力するようになっている。

#### 【0069】

なお、本実施形態における適応処理部330の構成およびデータ適応処理の詳細については、後述する。

#### 【0070】

雑音低減処理部350には、マイクロホン170を介してユーザが発話したマイク入力信号と、雑音推定処理部320において算出された雑音特徴量のデータ（以下、「雑音特徴量データ」という。）と、が入力されるようになっており、この雑音低減処理部350は、入力された雑音声特徴量データに基づいて、マイク入力信号全体の雑音低減処理をし、当該雑音低減処理が行われたマイク入力信号を音声信号分析部に出力するようになっている。

#### 【0071】

例えば、この雑音低減処理部350は、雑音が定常であることを利用して、雑音が含まれるマイク入力信号から当該雑音を取り除くスペクトルサブトラクション法を用いてマイク入力信号全体の雑音低減処理を行うようになっている。

#### 【0072】

音声分析部360には、雑音低減処理されたマイク入力信号が入力されるようになっており、この音声分析部360は、予め設定された時間間隔毎のフレームに分割するとともに、当該音声信号を分析し、フレーム毎の音声信号の特徴量を抽出して類似度算出部370に出力するようになっている。

#### 【0073】

具体的には、音声分析部360は、各フレーム毎に分割された音声信号に基づいて、当該各フレーム毎に、一定時間毎の各周波数毎におけるパワーを示すスペ

クトル包絡の情報、または、このパワースペクトルの対数をとて逆フーリエ変換したものケプストラムの情報を特徴量として抽出し、当該抽出した特徴量をベクトル化して類似度算出部370に出力ようになっている。

#### 【0074】

なお、例えば、音声分析部360は、1フレームを、10ms～20ms程度の時間間隔毎に分割するようになっている。

#### 【0075】

類似度算出部370には、各フレーム毎のベクトル特徴量とデータ適応処理されたキーワードモデルおよび不要語モデルが入力されるようになっており、この類似度算出部370は、入力された各フレーム毎の特徴量とデータ適応処理された各キーワードモデルおよび各不要語モデルの特徴量とを比較して、入力された各フレームの各キーワードモデルおよび各不要語モデルとの類似度を算出し、この算出された類似度をマッチング処理部390に出力するようになっている。

#### 【0076】

具体的には、類似度算出部370は、各フレーム毎の特徴量および各キーワードモデルおよび各不要語モデルの特徴量に基づいて、各フレームが各キーワードモデルおよび各不要語モデルを示す確率を算出するようになっている。

#### 【0077】

例えば、類似度算出部370は、各フレームが各キーワードモデルを示す出力確率、および、各フレームが不要語モデルを示す出力確率を算出するとともに、任意のフレームから次のフレームに状態遷移が各キーワードモデルから他のキーワードモデルまたは不要語モデルに状態遷移を示す状態遷移確率、および、不要語モデルから各キーワードモデルに状態遷移を示す状態遷移確率を算出し、これらの確率を類似度としてマッチング処理部390に出力するようになっている。

#### 【0078】

なお、状態遷移確率には、各キーワードモデルもしくは各不要語モデルの1つの状態から自己の状態への状態遷移を示す状態遷移確率、各キーワードモデルもしくは各不要語モデルの1つの状態から別の状態への状態遷移を示す状態遷移確率、キーワードモデルの最終状態から不要語モデルの初期状態への状態遷移を示

す状態遷移確率、および、不要語モデルの最終状態からキーワードモデルの初期状態への状態遷移を示す状態遷移確率が含まれるようになっている。

#### 【0079】

本実施形態では、類似度算出部370は、各フレーム毎に算出された各出力確率および各状態遷移確率を各フレームの類似度としてマッチング処理部390に出力するようになっている。

#### 【0080】

音声区間推定部380には、マイクロホン170を介してユーザが発話したマイク入力信号が入力されるようになっており、音声区間推定部380は、入力されたマイク入力信号に基づいて、音声区間の推定を行い、当該推定した結果を音声区間の情報としてマッチング処理部390に出力するようになっている。

#### 【0081】

マッチング処理部390には、各フレーム毎の各出力確率および各遷移出力確率と、音声区間推定部380によって推定された音声区間の情報が入力されるようになっており、マッチング処理部390は、この入力された音声区間の情報と、各フレーム毎の出力確率および遷移出力確率と、に基づいて各キーワードモデルと不要語モデルとの各組み合わせの類似度を示す累積類似度を算出するマッチング処理を行い、この算出された累積類似度を判定部400に出力するようになっている。

#### 【0082】

具体的には、マッチング処理部390は、音声区間に該当する各フレーム毎の各出力確率および各遷移出力確率を当てはめ、各キーワード毎に1の累積類似度を算出するようになっており、本実施形態では、マッチング処理部390は、ビタビアルゴリズムを用いて、キーワードモデルおよび不要語構成モデルとの各組み合わせの累積類似度を算出するようになっている。

#### 【0083】

なお、このビタビアルゴリズムは、各状態になる出力確率と各状態から他の状態へ遷移する場合の遷移確率に基づいて累積類似度を算出するアルゴリズムであり、任意の状態*i*から次の状態 *j*への遷移を示すパスに沿って計算される累積的

な確率の計算を行い、この累積的な確率計算を行うことによって状態の遷移が可能な各パス、すなわち、HMMの繋がりおよび組み合わせを抽出するようになっている。

#### 【0084】

また、本実施形態では、各キーワード毎に1の累積類似度を算出するようになっているが、キーワードなしの場合の累積類似度、すなわち、不要語構成要素モデルのみの累積類似度も算出するようになっている。

#### 【0085】

判定部400には、マッチング処理部390において算出された各キーワード毎の累積類似度および音声区間推定部380において推定された音声区間の情報が入力されるようになっており、判定部400は、推定された音声区間に合致し、かつ、入力された累積類似度の中で最も累積類似度が高いキーワードを、発話音声に含まれるキーワードと判定してこのキーワードをシステム制御部240に出力するようになっている。

#### 【0086】

このとき、この判定部400は、不要語構成要素モデルのみの累積類似度も判定対象に加えてキーワードの判定を行うようになっており、この不要語構成要素モデルのみの累積類似度が最も入力された累積類似度の中で最も高い場合には、発話音声にキーワードが含まれていなかったものと判定してこの判定結果をシステム制御部240に出力するようになっている。

#### 【0087】

次に、図3を用いて本実施形態における適応処理部の構成および動作について説明する。

#### 【0088】

なお、図3は、本実施形態における適応処理部の構成を示すブロック図である。

#### 【0089】

本実施形態の適応処理部330は、音声認識の処理を行う前に、キーワードモデルおよび不要語モデル毎に、クラスタリング処理、重心ベクトルの算出処理お

より差分ベクトルの算出処理を行うようになっている。

#### 【0090】

なお、HMMは、話者適応処理によって各HMMデータが変更されない限り、または、新規な話者によるHMMの話者適応が行われない限り、クラスタリング処理の結果は同じになる。したがって、本実施形態では、適応処理部330は、話者適応処理毎に、すなわち、HMMデータベース310に格納された各HMMデータに基づいて新たに話者適応されたHMMデータが生成される毎に、当該データ適応処理を行うようになっている。

#### 【0091】

本実施形態の適応処理部330は、図3に示すように、キーワードモデルおよび不要語モデル毎に、データ適応処理を行うようになっており、データ適応処理を行う不要語モデルおよびキーワードモデルに対してクラスタリング処理を行う第1クラスタリング処理部331および第2クラスタリング処理部341と、クラスタリング処理された不要語モデルまたはキーワードモデルに基づいて重心ベクトルを算出する第1重心ベクトル算出部332および第2重心ベクトル算出部342と、クラスタリング処理された不要語モデルまたはキーワードモデルと算出された重心ベクトルに基づいて差分ベクトルを算出する第1差分ベクトル算出部333および第2差分ベクトル算出部343と、算出された重心ベクトルと雑音モデルに基づいて各HMMに対してデータ適応処理を行う第1モデル適応処理部334および第2モデル適応処理部344と、データ適応処理された各HMMと算出された差分ベクトルに基づいて各HMMの復元処理を行う第1復元処理部335および第2復元処理部345と、から構成されている。

#### 【0092】

なお、例えば、本実施形態の第1クラスタリング処理部331および第2クラスタリング処理部341は、本発明の分類手段を構成し、第1重心ベクトル算出部332および第2重心ベクトル算出部342は、本発明の抽出手段を構成する。

#### 【0093】

また、例えば、本実施形態の第1差分ベクトル算出部333および第2差分ベ

クトル算出部343は、本発明の第1算出手段を構成し、第1モデル適応処理部334および第2モデル適応処理部344は、本発明の第2算出手段を構成する。

#### 【0094】

また、例えば、本実施形態の第1復元処理部335および第2復元処理部345は、本発明の音声特徴量データ生成手段を構成する。

#### 【0095】

第1クラスタリング処理部331には、HMMデータベース310に格納されている全ての不要語モデルであるHMMが入力されるようになっており、この第1クラスタリング処理部331は、入力された全てのHMMに対してクラスタリング処理を行い、その処理結果を示す各HMMの属性データを当該各HMMとともに、重心ベクトル算出部および差分ベクトル算出部に出力するようになっている。

#### 【0096】

具体的には、第1クラスタリング処理部331は、図4に示すように、入力された全ての不要語モデルである全てのHMMを1つの集合として、当該集合（以下、「クラスタ」という。）を、予め定められた規則に従って所定数の複数のHMMの集合群に分割するようになっており、各HMMがこの分割された各集合群に属することを示す属性データを、当該各HMMとともに、重心ベクトル算出部および差分ベクトル算出部に出力するようになっている。

#### 【0097】

本実施形態の第1クラスタリング処理部331は、HMMのクラスタに対して、初回を除き、分割されて生成された各クラスタの歪みのうち、最大となるクラスタを2つに分割し、目的の集合数になるまで、この分割を繰り返すアンバランストクラスタリングを用いて当該HMMのクラスタリング処理を行うようになっている。

#### 【0098】

具体的には、本実施形態では、第1クラスタリング処理部331は、各クラスタに属する各HMMに対して、それぞれ、各クラスタの重心との差分ベクトルを

算出するようになっており、各クラスタにおいて最大となるHMMとの差分ベクトルの大きさをそのクラスタの最大歪みとして算出するようになっている。

#### 【0099】

なお、各HMM毎に示される属性データとは、各HMMが属するクラスタの識別符号のデータをいい、例えば、クラスタの識別符号が番号の場合には、その番号を示す。

#### 【0100】

第1重心ベクトル算出部332には、第1クラスタリング処理部331において算出された各HMMの属性データと、当該各HMMと、が入力されるようになっており、この第1重心ベクトル算出部332は、入力された各属性データと各HMMとに基づいて各クラスタの重心を算出し、当該算出した各クラスタの重心ベクトルをベクトルデータ（以下、「重心ベクトルデータ」という。）として第1モデル適応処理部334および第1差分ベクトル算出部333に出力するようになっている。

#### 【0101】

具体的には、第1重心ベクトル算出部332は、入力された各クラスタのHMMに基づいて、各HMMのベクトルの平均を算出し、この平均されたベクトルを重心ベクトルとして算出するようになっている。

#### 【0102】

第1差分ベクトル算出部333には、第1クラスタリング処理部331において算出された各HMMの属性データおよび各HMMと、第1重心ベクトル算出部332において算出された各クラスタの重心ベクトルデータが入力されるようになっている。

#### 【0103】

この第1差分ベクトル算出部333は、入力された各属性データ、そのHMMおよび当該HMMが属するクラスタの重心ベクトルデータに基づいて、各HMM毎に、重心ベクトルデータとの差分ベクトルを算出し、当該算出した各HMMの差分ベクトルを、差分ベクトルデータとして各HMMの属性データに対応付けて第1復元処理部335に出力するようになっている。

### 【0104】

第1モデル適応処理部334には、第1重心ベクトル算出部332において算出された各クラスタの重心ベクトルデータと適応パラメータとしての雑音モデルが入力されるようになっており、第1モデル適応処理部334は、入力された雑音モデルに基づいて各クラスタの重心ベクトルに対してモデル適応を行い、当該モデル適応した各クラスタの重心ベクトルを第1復元処理部335に出力するようになっている。

### 【0105】

具体的には、第1モデル適応処理部334は、各クラスタの重心ベクトルに対して雑音モデルを合成するようになっており、各重心ベクトルに雑音モデルを重畳し、各重心ベクトルに適応処理（以下、「モデル適応処理」という。）を行うようになっている。

### 【0106】

なお、本実施形態では、各クラスタの重心ベクトルに対して雑音モデルを合成する方法としては、公知のHMM合成法を用いるようになっており、例えば、PMC法（Parallel Model Combination）やNOVO合成法（Voice Mixed with Noise）を用いるようになっている。

### 【0107】

第1復元処理部335には、第1差分ベクトル算出部333において算出された各HMMの差分ベクトルおよびその属性データと、第1モデル適応処理部334においてモデル適応された各クラスタの重心ベクトルが入力されるようになっている。

### 【0108】

この第1復元処理部335は、入力された各クラスタの重心ベクトルと、各HMMの差分ベクトルおよびその属性データに基づいて各HMMの復元の処理（以下、単に「復元処理」という。）を行い、当該復元した各HMMを適応処理部330の出力として類似度算出部370に出力するようになっている。

### 【0109】

具体的には、第1復元処理部335は、属性データに基づいて、上述の差分ベ

クトルの算出処理と反対の処理、すなわち、適応処理された各重心ベクトルに、各HMMモデルの差分ベクトルを加算して各HMMを復元するようになっている。

#### 【0110】

なお、本実施形態の適応処理部330は、上述のように、不要語モデルに対してデータ適応処理を行う各部について説明したが、図3に示すように、キーワードモデルに対しても不要語モデルと同様に、データ適応処理を行うようになっている。

#### 【0111】

すなわち、本実施形態では、第2クラスタリング処理部341、第2重心ベクトル算出部342、第2差分ベクトル算出部343、第2モデル適応処理部344および第2復元処理部345によって、HMMデータベース310に格納された各キーワードモデルに対して、不要語モデル同様に、クラスタリング処理、重心ベクトルの算出処理（以下、「重心ベクトル算出処理」という。）、差分ベクトルの算出（以下、「差分ベクトル算出処理」という。）、モデル適応処理、および復元処理を行い、復元処理した各キーワードのHMMを適応処理部330の出力として類似度算出部370に出力するようになっている。

#### 【0112】

また、本実施形態では、適応処理部330は、クラスタリング処理、重心ベクトル算出処理および差分ベクトル算出処理を各HMM毎に、かつ、話者毎に、予め実行するようになっており、後述する音声認識処理を行うときに、予め算出してある各HMMの重心ベクトル、差分ベクトルおよび属性データに基づいて、話者毎に適応処理したHMMを類似度算出部370に出力するようになっている。

#### 【0113】

次に、図4および図5を用いて、不要語モデルおよびキーワードモデル毎にクラスタリング処理、および重心ベクトルの算出処理の原理について説明する。

#### 【0114】

なお、図4は、本実施形態における重心ベクトルの算出を説明するための図であり、図5は、実施形態のクラスタリング処理を説明するための図である。

### 【0115】

図4に示すように、複数の不要語モデルおよびキーワードが存在すると、不要語モデルのみからなる重心位置およびキーワードモデルのみからなら重心位置と、キーワードモデルおよび不要語モデルが混在する重心位置は、異なる位置となる。

### 【0116】

すなわち、本実施形態では、重心ベクトルに基づいて各不要語モデルおよびキーワードモデルの復元処理を行うため、重心ベクトルの位置に基づいて雑音モデルに適応処理された各不要語モデルおよびキーワードモデルに基づいて音声認識が行われるので、この各クラスタの重心を的確に算出する必要がある。

### 【0117】

一方、不要語モデルとキーワードモデルでは、生成されるHMMモデルの特徴が異なる。すなわち、キーワードモデルは、その性質上、該当する1のキーワードの音声特徴量を現せば十分であるが、不要語モデルは、複数の不要語を識別させるため、1のHMMモデルで複数の不要語の音声特徴量を表すようになっている。

### 【0118】

したがって、キーワードモデルと不要語モデルは、HMMモデル自体の性質が異なるため、クラスタにキーワードモデルおよび不要語モデルが混在すると、不要語モデルのみからなるクラスタおよびキーワードモデルのみのクラスタにクラスタリング処理することができず、不要語モデルのみからなるクラスタおよびキーワードモデルのみのクラスタに比べ、これらが混在するクラスタから算出される重心位置が的確な位置を示さないこととなる。

### 【0119】

そこで、本実施形態は、不要語モデルまたはキーワードモデル毎にクラスタリング処理を行うようになっており、不要語モデルまたはキーワードモデル毎にクラスタリング処理された集合群（クラスタ）に基づいて、重心ベクトルおよびこの重心ベクトルとの差分ベクトルを算出するようになっている。

### 【0120】

また、クラスタリング処理とは、不要語モデルまたはキーワードに基づいて予め定められた規則に基づいて予め定められた分割数になるまで分割するとともに、分割したクラスタに対して重心ベクトルを算出し、当該重心ベクトルに基づいて各HMMの復元処理を行うようになっており、各HMM毎にモデル適応処理を行う場合に比べ、処理負担の軽減を行うようになっている。すなわち、分割されたクラスタの数は、HMMの格納されている数に比べ、少なくなるので、この少なくなった分の処理が軽減されるようになっている。

#### 【0121】

なお、本実施形態では、例えば、図5に示すように、不要語モデルに対して3回の分割を行う場合には、第1クラスタリング処理部は、まず、全ての不要語モデルにおける集合（クラスタ）を2つのクラスタに分割し、次いで、2つの分割されたクラスタのうち、歪みが最大となるクラスタを選択し、その選択したクラスタをさらに2つに分割するようになっており、次いで、分割されていないクラスタと、2つに分割された各クラスタと、において、最大の歪みとなるクラスタをさらに分割し、目的の集合数、すなわち、分割数になるまでこの分割処理を行うようになっている。

#### 【0122】

次に、図6を用いて本実施形態における音声認識処理の動作について説明する。

#### 【0123】

なお、図6は、本実施形態のナビゲーション装置100における音声認識処理の動作を示すフローチャートである。

#### 【0124】

まず、操作部160または音声認識部300を介してシステム制御部240にユーザから音声認識処理の指示が入力されると、または、ナビゲーション処理中に、システム制御部240が音声認識処理を行う必要があると判断すると（ステップS11）、システム制御部240は、音声処理回路210およびスピーカ220を介して、音声認識すべき発話音声をマイクロホン170から入力する旨をユーザに告知する（ステップS12）。

**【0125】**

次いで、ユーザが発話し、発話を含む信号が入力されると、システム制御部240は、音声認識部300に、入力された発話音声に対して以下の音声認識処理を行わせる。

**【0126】**

まず、発話音声が入力されると、雑音推定処理部320は、予め設定された初期区間の初期入力信号に基づいて雑音モデルを算出するとともに、雑音特徴量を算出し、雑音低減処理部350および適応処理部330に出力する（ステップS13）。

**【0127】**

次いで、適応処理部330に雑音モデルが入力されると、当該適応処理部330は、予めクラスタリング処理して算出した各キーワードモデルおよび不要語モデルの重心ベクトル、差分ベクトルおよびその属性データと、入力された雑音性モデルに基づいて、上述のようにモデル適応処理を行うとともに、復元処理を行い、復元した各キーワードモデルおよび各不要語モデルを類似度算出部370に出力する（ステップS14）。

**【0128】**

なお、この適応処理部330におけるクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理の動作については、後述する。

**【0129】**

一方、雑音低減処理部350にマイク入力信号が入力され、かつ、雑音特徴量が入力されると、当該雑音処理部は、入力された雑音声特徴量データに基づいて、マイク入力信号全体の雑音低減処理をし、当該雑音低減処理が行われたマイク入力信号を音声分析部360に出力する（ステップS15）。

**【0130】**

次いで、音声分析部360に雑音低減処理されたマイク入力信号が入力されると、当該音声分析部360は、入力されたマイク入力信号の発話音声部分の音声区間を示す音声信号をフレームに分割するとともに、当該各フレーム毎の音声信号の特徴量を抽出して類似度算出部370に出力する（ステップS16）。

**【0131】**

次いで、類似度算出部370に各適応処理されたキーワードモデルおよび不要語モデルが入力され、かつ、各フレーム毎のベクトル特微量が入力されると、当該類似度算出部370は、入力された各フレーム毎の特微量とHMMデータベース310に格納されている各HMMの特微量とを比較して、入力された各フレームの各HMMとの類似度を算出し、この算出された類似度をマッチング処理部390に出力する（ステップS17）。

**【0132】**

一方、音声区間推定部380にマイク入力信号が入力されると、音声区間推定部380は、マイク入力信号に基づいて、音声区間の推定を行い、当該推定した音声区間の情報をマッチング処理部390に出力する（ステップS18）。

**【0133】**

次いで、マッチング処理部390に各フレーム毎の各出力確率および各遷移出力確率と、音声区間推定部380によって推定された音声区間の情報が入力されると、当該マッチング処理部390は、この入力された音声区間の情報と、各フレーム毎の出力確率および遷移出力確率と、に基づいて各キーワードモデルと不要語モデルとの各組み合わせの類似度を示す累積類似度を算出するマッチング処理を行い、この算出された累積類似度を判定部400に出力する（ステップS19）。

**【0134】**

次いで、判定部400にマッチング処理部390において算出された各キーワード毎の累積類似度および音声区間推定部380において推定された音声区間の情報が入力されると、当該判定部400は、推定された音声区間において、入力された累積類似度の中で最も累積類似度が高いキーワードを、発話音声に含まれるキーワードと判定してこのキーワードをシステム制御部240に出力する（ステップS20）。

**【0135】**

最後に、システム制御部240は、判定部400から出力されたキーワードに基づいて各部を制御し（ステップS21）、本動作を終了する。

**【0136】**

次に、図7を用いて、適応処理部330におけるクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理の動作について説明する。

**【0137】**

なお、図7は、本実施形態の適応処理部330におけるクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理の動作を示すフローチャートである。

**【0138】**

また、本動作は、操作部160を介してシステム制御部240に話者が変わった旨または話者の新規登録を行う旨が入力され、各キーワードモデルおよび各不要語モデルに対して当該話者適応が行われた後に、システム制御部240の指示に基づいて行われるようになっている。

**【0139】**

まず、システム制御部240が音声認識部300にクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理を実行する旨が入力されると、システム制御部240は、第1クラスタリング処理部331および第2クラスタリング処理部341に、HMMデータベース310に格納された全てのキーワードモデルまたは全ての不要語モデルを取得させる（ステップS31）。

**【0140】**

次いで、システム制御部240は、各キーワードモデル毎に、または、各不要語モデル毎に、以下の動作を行う。

**【0141】**

まず、システム制御部240は、第1クラスタリング処理部331に、全ての不要語モデルに基づいてクラスタリング処理を行わせ、各不要語モデル毎に属性データを算出させる（ステップS32）。

**【0142】**

次いで、システム制御部240は、第1重心ベクトル算出部332に、各不要語モデル毎に、各不要語モデルおよびその属性データに基づいて重心ベクトルを算出させ、第1モデル適応処理部334に出力する（ステップS33）。

**【0143】**

次いで、システム制御部240は、第1差分ベクトル算出部333に、各不要語モデル毎に、各キーワードモデル、その属性データおよびその重心ベクトルに基づいて差分ベクトルを算出させる（ステップS34）。

**【0144】**

一方、システム制御部240は、第2クラスタリング処理部341に、全てのキーワードモデルに基づいてクラスタリング処理を行わせ、各キーワードモデル毎に属性データを算出させる（ステップS35）。

**【0145】**

次いで、システム制御部240は、第2重心ベクトル算出部342に、各キーワードモデル毎に、各キーワードモデルおよびその属性データに基づいて重心ベクトルを算出させる（ステップS36）。

**【0146】**

次いで、システム制御部240は、第2差分ベクトル算出部343に、各キーワードモデル毎に、各キーワードモデル、その属性データおよびその重心ベクトルに基づいて差分ベクトルを算出させる（ステップS37）。

**【0147】**

最後に、システム制御部240は、算出した各キーワードモデルおよび不要語モデルの重心ベクトル、差分ベクトルおよび属性データを、それぞれ、第1モデル適応処理部334、第2モデル適応処理部344、第1復元処理部335および第2復元処理部345に出力させる（ステップS38）。

**【0148】**

なお、本実施形態では、不要語モデルにおいてクラスタリング処理を行うとともに、重心ベクトルおよび差分ベクトルを算出した後に、キーワードモデルにおいてクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理を行っているが、不要語モデルよりもキーワードモデルにおいてクラスタリング処理等を行った後に、不要語モデルのクラスタリング処理を行うようにしてもよく、また、不要語モデルおよびキーワードモデルのクラスタリング処理を同時並行に行うようにしてもよい。

**【0149】**

次に、図8を用いて、本実施形態における音声認識処理の認識率の実験結果について説明する。

**【0150】**

なお、図8は、本実施形態における音声認識処理の認識率と不要語モデルおよびキーワードモデル毎にクラスタリング処理を行わない従来の方法における音声認識処理の認識率を示す表である。また、この図8は、自動車内で発話された音声に対する認識率を示している。

**【0151】**

ケース1は、キーワードだけが発話された場合の認識率であり、ナビゲーション装置が搭載された車両の時速が、0 km/h、50 km/h および100 km/h の場合に、男性11名、女性9名が発声したそれぞれ150語のナビゲーションの指示命令を認識させ、その結果を平均したものである。

**【0152】**

ケース2は、キーワードのワード長よりも短い不要語が含まれる発話音声が発話された場合の認識率であり、ナビゲーション装置が搭載された車両の時速が、0 km/h、50 km/h および100 km/h の場合に、男性11名、女性9名が発声したそれぞれ150語のナビゲーションの指示命令を認識させ、その結果を平均したものである。

**【0153】**

ケース3は、キーワードのワード長と同程度かそれ以上のワード長を持つ不要語が含まれる発話音声が発話された場合の認識率であり、ナビゲーション装置が搭載された車両の時速が、100 km/h の場合に、男性5名、女性4名が発声したそれぞれ200語のナビゲーションの指示命令を認識させ、その結果を平均したものである。

**【0154】**

このように、本実施形態の音声認識処理は、どのケースにおいても発話音声の認識率の劣化を従来のナビゲーション装置よりも小幅に留めつつ、当該認識率を向上させていることがわかる。

### 【0155】

以上により、本実施形態によれば、複数のHMMが予め格納されているHMMデータベース310と、格納されているHMMの種別毎に、予め設定された規則に基づいて、当該HMMを複数のクラスタにクラスタリング処理を行う第1クラスタリング処理部331および第2クラスタリング処理部341と、クラスタリング処理された各クラスタ毎に、クラスタの特徴量を示す重心ベクトルを抽出する第1重心ベクトル算出部332および第2重心ベクトル算出部342と、発話音声が発せられた環境の状況を示す雑音環境を示す雑音モデルを取得する雑音環境を示す雑音推定処理部320と、格納されている各HMM、クラスタリング処理された各クラスタの属性を示す属性データ、取得された各重心ベクトル、および、雑音環境を示す雑音モデルに基づいて、発話音声の特徴量に照合する際に用いられるHMMを生成する適応処理部330と、生成された各HMMと発話音声の特徴量とを照合することによって発話音声を認識する類似度算出部370、マッチング処理部390および判定部400と、を備えた構成を有している。

### 【0156】

この構成により、本実施形態では、格納されている各HMM、クラスタリング処理された各クラスタの属性を示す属性データ、取得された各重心ベクトル、および、雑音環境を示す雑音モデルに基づいて、発話音声の特徴量に照合する際に用いられるHMMを生成するとともに、生成された各HMMと発話音声の特徴量とを照合することによって発話音声を認識する。

### 【0157】

したがって、キーワードモデルおよび不要語モデル毎に、クラスタリング処理を行えば、キーワードモデルおよび不要語モデル毎に、発話音声が発話された雑音環境に適応処理されたHMMモデルを生成することができるので、不要語モデルおよびキーワードモデル間の音声特徴量の大幅な特性の差異によって生ずる誤認識を防止することができ、高精度に音声認識を行うことができる。

### 【0158】

また、本実施形態では、適応処理部330が、各HMMと当該各HMMが属するクラスタにおける重心ベクトルとの差異を示す差分ベクトルを算出する第1差

分ベクトル算出部333および第2差分ベクトル算出部343と、取得された各重心ベクトルに雑音モデルを、それぞれ、重畳して発話環境に適応した重心ベクトルである適応重心ベクトルを算出する第1モデル適応処理部334および第2モデル適応処理部344と、算出された各HMMにおける差分ベクトルおよび属性データと算出された各適応重心ベクトルとに基づいて発話音声の特徴量に照合する際に用いられるHMMを復元する第1復元処理部335および第2復元処理部345と、を有する構成をしている。

#### 【0159】

この構成により、本実施形態では、取得された各重心ベクトルに雑音モデルを、それぞれ、重畳して適応重心ベクトルを算出するとともに、各HMMにおける算出された差分ベクトルおよび属性データと、算出された各適応重心ベクトルと、に基づいて発話音声の特徴量に照合する際に用いられるHMMを復元する。

#### 【0160】

したがって、各クラスタ毎の重心ベクトルおよび各HMM毎の差分ベクトルを用いることによって、各重心ベクトル毎に適応処理を行えば、雑音環境に適応したHMMを生成することができるので、各HMM毎に適応処理を行うことに比べ、適応処理を迅速に行うことができる。

#### 【0161】

また、本実施形態では、各重心ベクトル算出部、および、各差分ベクトル算出部が、それぞれ、重心ベクトルおよび差分ベクトルを予め算出する構成を有している。

#### 【0162】

この構成により、本実施形態では、予め各クラスタの重心ベクトルおよび各HMMの差分ベクトルを算出することができるので、キーワードモデルおよび不要語モデル毎に、クラスタリング処理を行うことができる。したがって、不要語モデルおよびキーワードモデル間の音声特徴量の大幅な特性の差異によって生ずる誤認識を防止することができ、迅速に、かつ、高精度に音声認識を行うことができる。

#### 【0163】

また、本実施形態では、発話された発話音声の特徴量を認識すべきキーワードと当該キーワードを構成しない不要語とに識別することによって音声認識を行う場合に、HMMデータベース310にキーワードのHMMおよび不要語のHMMが格納されているとともに、第1クラスタリング処理部331および第2クラスタリング処理部341が、キーワードおよび不要語のHMMの種別毎に、予め設定された規則に基づいて、当該HMMを複数のクラスタにクラスタリング処理を行う構成を有している。

#### 【0164】

この構成により、本実施形態では、キーワードおよび不要語のHMMの種別毎に、予め設定された規則に基づいて、当該HMMを複数のクラスタにクラスタリング処理する。

#### 【0165】

したがって、キーワードモデルおよび不要語モデル毎に、発話音声が発話された雑音環境に適応処理されたHMMモデルを生成することができるので、不要語モデルおよびキーワードモデル間の音声特徴量の大幅な特性の差異によって生ずる誤認識を防止することができ、迅速に、かつ、高精度に音声認識を行うことができる。

#### 【0166】

また、本実施形態では、発話音声を分析することによって当該発話音声の音声成分の特徴量である発話音声特徴量を取得する音声分析部360と、マイク入力信号の少なくとも一部の区間の取得された発話音声特徴量と適応処理されたHMMとを照合し、特徴量の特性が類似する割合を示す類似度を算出する類似度算出手段と、算出した類似度に基づいて発話音声を認識するマッチング処理部390および判定部400と、を有する構成をしている。

#### 【0167】

この構成により、本実施形態では、マイク入力信号の少なくとも一部の音声区間の取得された発話音声特徴量と適応処理されたHMMとを照合し、特徴量の特性が類似する割合を示す類似度を算出するとともに、算出した類似度に基づいて発話音声を認識する。

**【0168】**

したがって、キーワードモデルおよび不要語モデル毎に、発話音声が発話された雑音環境に適応処理されたHMMモデルを生成することができるので、不要語モデルおよびキーワードモデル間の音声特徴量の大幅な特性の差異によって生ずる誤認識を防止することができ、高精度に音声認識を行うことができる。

**【0169】**

なお、本実施形態では、適応処理部330は、音声認識処理を行う前に、クラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理を行うようになっているが、音声認識処理毎に、または、操作部160によるユーザの指示に基づいて、クラスタリング処理、重心ベクトル算出処理および差分ベクトル算出処理を行うようにしてもよい。

**【0170】**

また、本実施形態では、上述のナビゲーション装置100によって、音声認識処理を行うようになっているが、認識すべき発話音声を入力するマイクロホン170を有するナビゲーション装置100にコンピュータおよび記録媒体を備え、この記録媒体に上述のデータ適用処理を含む音声認識処理を行うプログラムを格納し、このコンピュータで当該プログラムを読み込むことによって上述と同様の音声認識処理を行うようにしてもよい。

**【図面の簡単な説明】****【図1】**

本願に係るナビゲーション装置の概要構成を示すブロック図である。

**【図2】**

一実施形態における音声認識部の構成を示すブロック図である。

**【図3】**

一実施形態における適応処理部の構成を示すブロック図である。

**【図4】**

一実施形態における重心ベクトルの算出を説明するための図である。

**【図5】**

一実施形態のクラスタリング処理を説明するための図である。

**【図6】**

一実施形態のナビゲーション装置における音声認識処理の動作を示すフローチャートである。

**【図7】**

一実施形態の適応処理部におけるクラスタリング処理、重心ベクトルの算出処理および差分ベクトルの算出処理の動作を示すフローチャートである。

**【図8】**

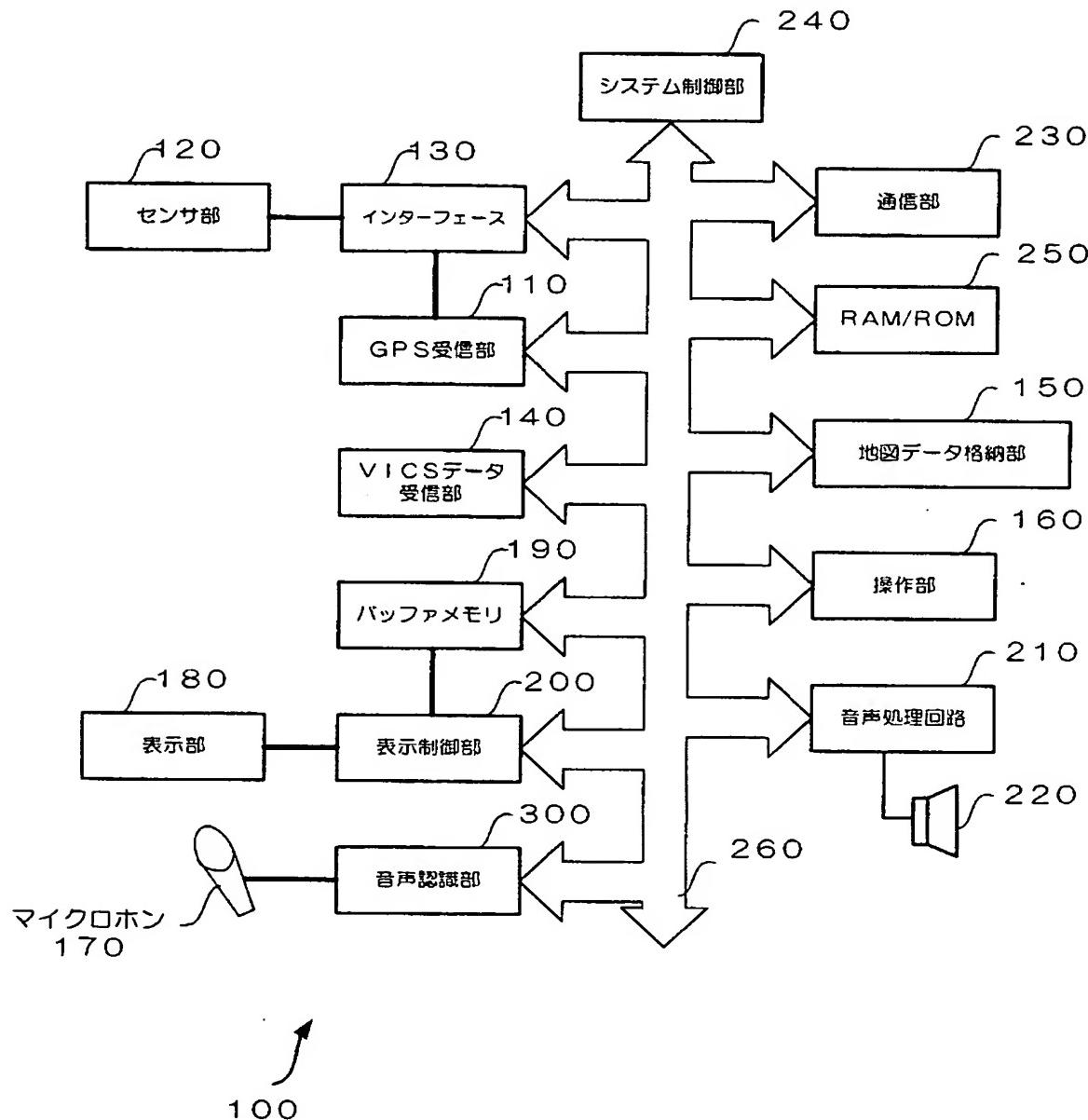
一実施形態のナビゲーション装置と従来のナビゲーション装置の音声認識率を比較した表である。

**【符号の説明】**

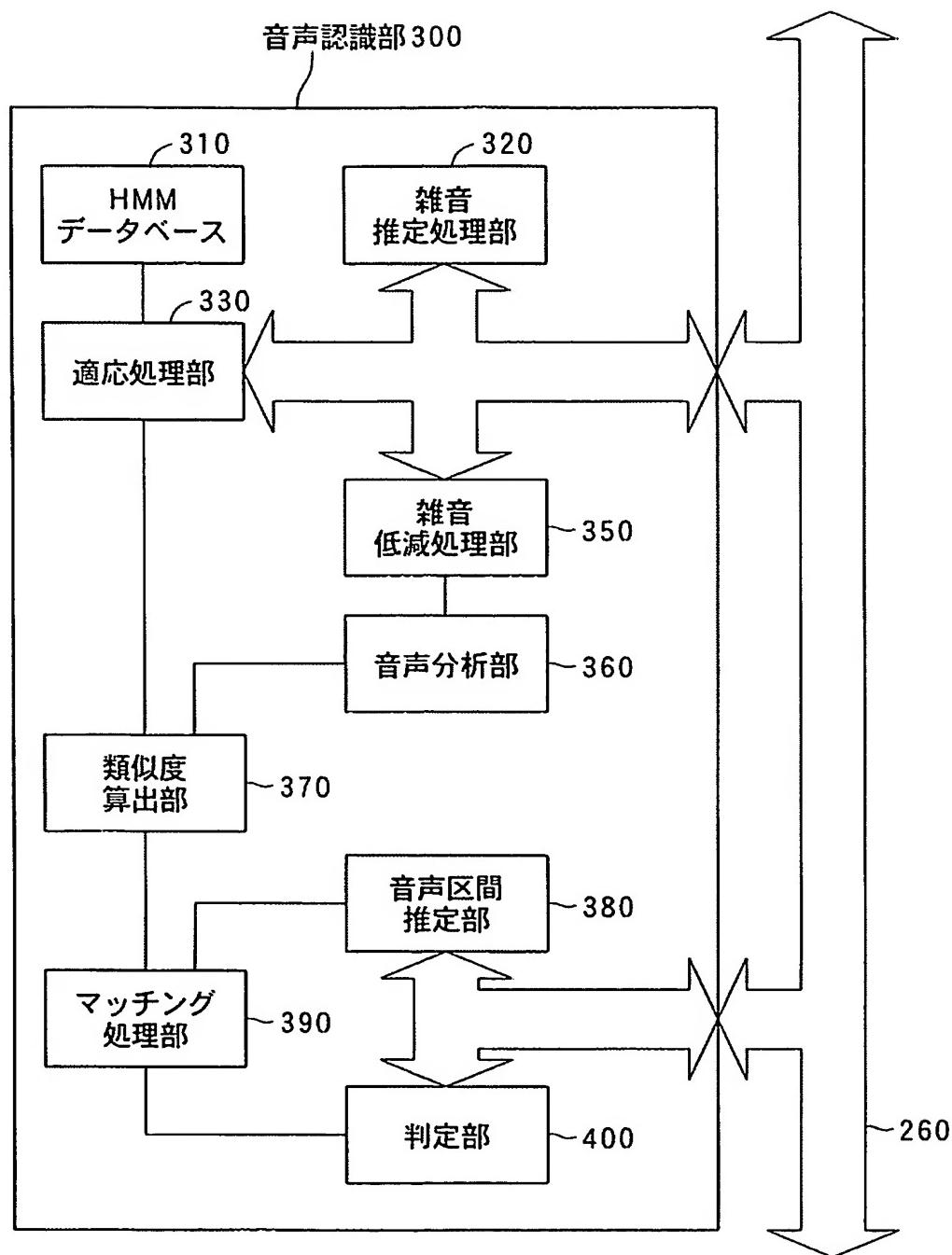
- 100 … ナビゲーション装置
- 300 … 音声認識部
- 310 … HMMデータベース
- 320 … 雑音推定処理部
- 330 … 適応処理部
- 331 … 第1クラスタリング処理部
- 332 … 第1重心ベクトル算出部
- 333 … 第1差分ベクトル算出部
- 334 … 第1モデル適応部
- 335 … 第1復元処理部
- 341 … 第2クラスタリング処理部
- 342 … 第2重心ベクトル算出部
- 343 … 第2差分ベクトル算出部
- 344 … 第2モデル適応部
- 345 … 第2復元処理部
- 370 … 類似度算出部
- 390 … マッチング処理部
- 400 … 判定部

【書類名】 図面

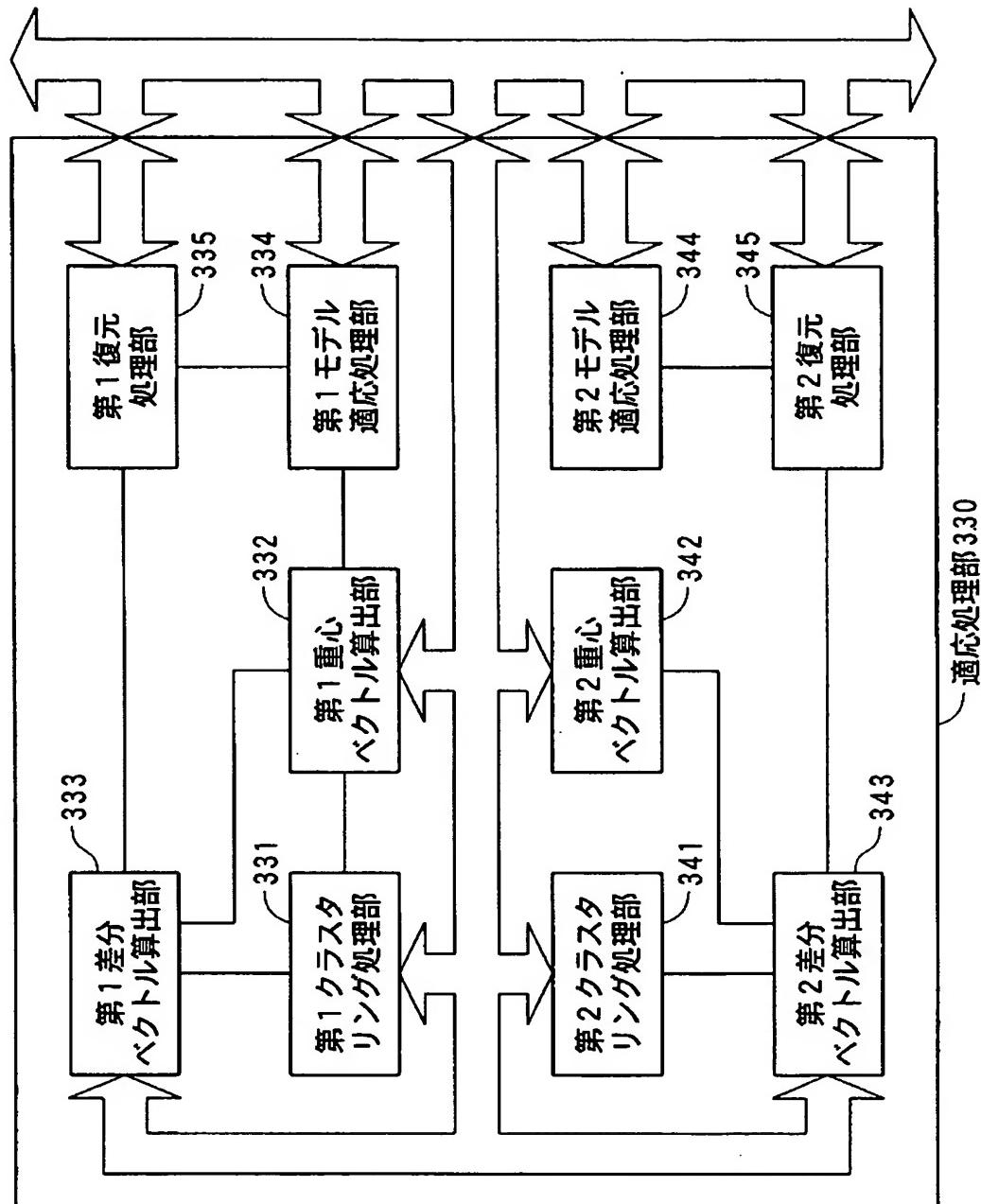
【図1】



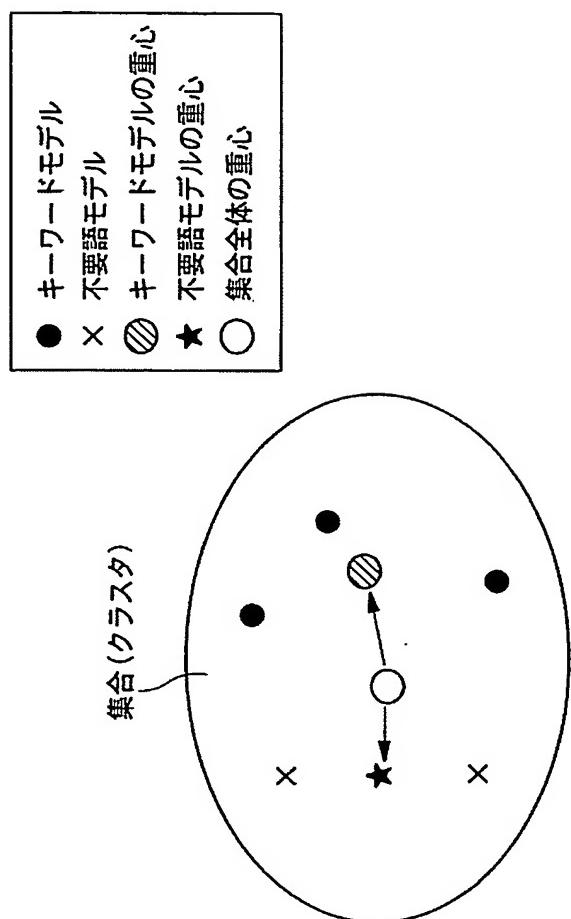
【図2】



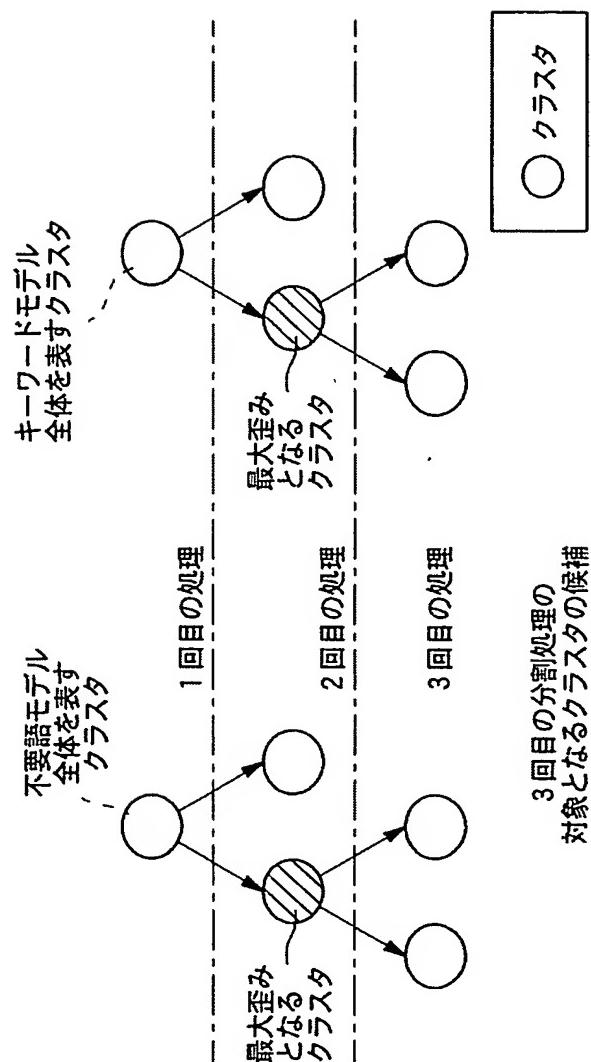
【図3】



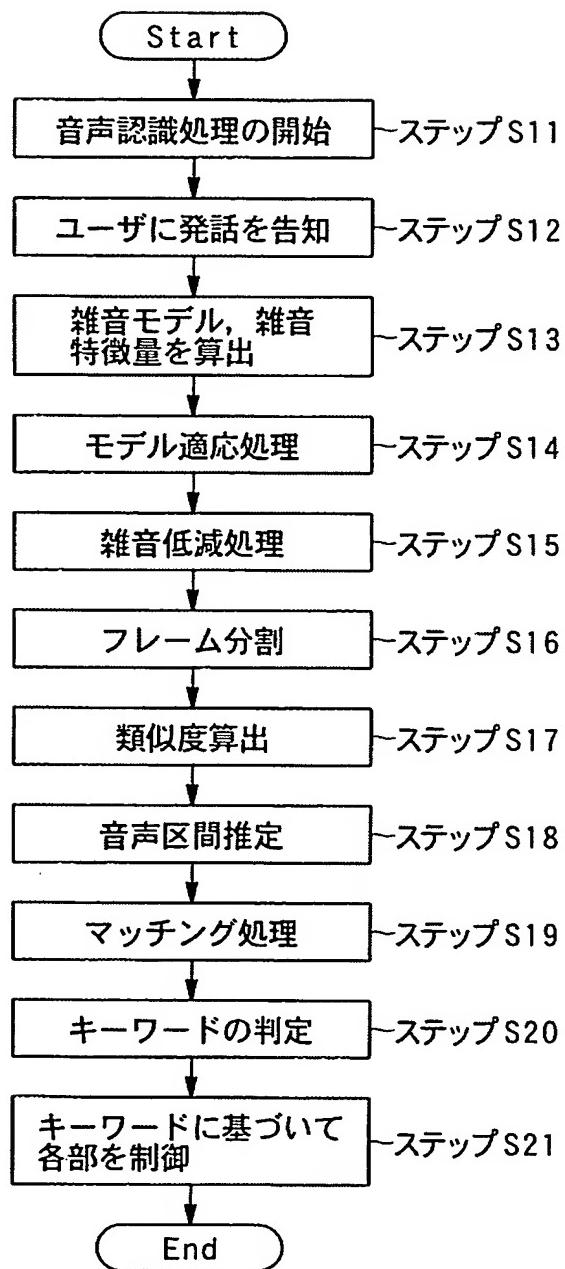
【図4】



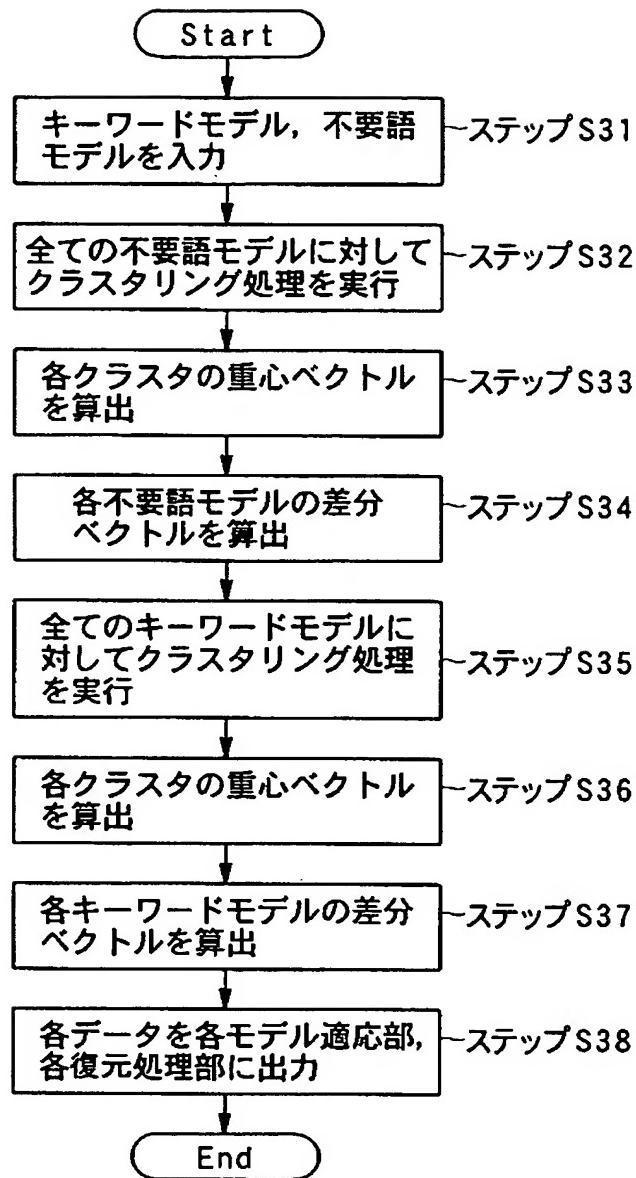
【図 5】



【図6】



【図7】



【図8】

	ケース1(%)	ケース2(%)	ケース3(%)	平均(%)
従来	93.3	69.2	54.1	72.2
本実施形態	94.1	68.9	58.5	73.8

【書類名】 要約書

【要約】

【課題】 キーワードモデルおよび不要語モデル毎に、認識を実行する条件に基づいてこれらのモデルに対して適応処理を行い、音声認識の性能が高く、データ処理を迅速に行うことができる音声認識装置を提供すること。

【解決手段】 本実施形態のナビゲーション装置 100 に組み込まれる音声認識部 300 は、入力されたマイク入力信号に基づいて雑音モデルを算出する雑音推定処理部 320 と、雑音モデルに基づいて HMM データベース 310 に格納されているキーワードモデルおよび不要語モデル毎に、当該各モデルに対して適応処理を行う適応処理部 330 と、を有し、適応処理部 330 が、キーワードモデルおよび不要語モデル毎に、雑音モデルに基づいてデータ適応処理を行うようになっており、このデータ適用処理が為されたキーワードモデルおよび不要語モデルに基づいてワードスポットティングを行うようになっている。

【選択図】 図 3

特願2003-122337

出願人履歴情報

識別番号 [000005016]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住所 東京都目黒区目黒1丁目4番1号  
氏名 パイオニア株式会社